

REGULAR ARTICLE

Open Access



Cities of a feather flock together: a study on the synchronization of communication between Italian cities

Lorenzo Candeago^{1,2,3*} , Giulia Bertagnolli^{1,4}, Paolo Bosetti^{1,4}, Michele Vescovi², Francesco Sacco⁵ and Bruno Lepri¹

*Correspondence:
candeago@fbk.eu

¹Fondazione Bruno Kessler, Trento, Italy

²SKIL (Semantic and Knowledge Innovation Lab) Joint Open Lab, TIM—Telecom Italia, Trento, Italy

Full list of author information is available at the end of the article

Abstract

Due to the rise of communication technologies and economic globalization, modern large cities are becoming more and more interconnected and this phenomenon leads to an increasing synchronization in activities and communication patterns. In our work, we explore the communication synchronization between 76 Italian cities of different sizes by using mobile phone data. Our results show that both the spatial distance and the size of the city influence the synchronization: larger cities are more similar to larger cities in communication rhythms than medium cities are to medium cities, and medium cities are more similar to medium cities than smaller cities are to smaller cities. Furthermore, for all the cities' sizes we observe a drift in similarity due to spatial distance. Interestingly, the drift due to distance over similarity is less strong in large cities, that act as gateway nodes for the Italian economical system, hence having an emerging strongly connected and synchronized network, than for medium and small cities, that are more bounded to local industries. Finally, our results also show that highly synchronized cities are richer and more attractive for foreign-born population.

Keywords: Communication synchronization; Mobile phone data; Computational social science

1 Introduction

Synchronization is a spontaneous process that emerges in many domains in nature [1], from neurons [2], trees [3], animals [4], and up to human beings [5–7].

Nowadays, modern large cities are becoming more and more interconnected [8] and this phenomenon leads to an increasing communication and activities' synchronization, as observed in Morales *et al.* [9]. Recent urban sociology literature [10, 11] has investigated the synchronization due to globalization, where large companies tend to spread their headquarters are in different cities and countries. This literature [10, 11] has also introduced the concept of *gateway cities*, namely central nodes for all the communication and economical activities, and for the people flows from and to the region where the city is located.

In our work, we explore the communication synchronization between 76 Italian cities of different sizes by using mobile phone data (i.e., Call Detail Records) and investigate

© The Author(s) 2019. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

if and which Italian cities act as *gateway cities*. We also explored how the synchronization between couples of cities changes depending on the size of the city and the spatial distance between them. We found that both spatial distance and city size influence the synchronization: larger cities are more similar to larger cities in communication rhythms than medium cities are to medium cities, and medium cities are more similar to medium cities than smaller cities are to smaller cities. Moreover, for all the cities' sizes we observed a drift in similarity due to spatial distance. In addition, we have also investigated if cities with a higher average synchronization tend to be richer and to attract more people from other places. Our results show that highly synchronized cities have a higher percentage of foreign-born population and higher levels of average yearly income per tax payer.

The remainder of the paper is structured as follows: in Sect. 2 we describe the mobile phone data and the socio-economic indicators (Sect. 2.1), we introduce the Dynamic Time Warping distance algorithm (Sect. 2.2) that we use for computing the synchronization among the communication activity timeseries of our cities, and finally we describe the bootstrap resampling procedure used (Sect. 2.3). In Sect. 3 we present our results on communication synchronization and on the influence played by the city size and the spatial distances as well as the associations between the synchronization of calling patterns of a city and the city's socio-economic indicators. Moreover, in Sect. 4 we discuss the obtained results with regard to the urban sociology literature and expose the limitations of our approach. Finally, in Sect. 5 we draw our conclusions.

2 Materials and methods

2.1 Data description

Our dataset consists of 24 consecutive days (18 weekdays and 6 weekend days) of Call Detail Records (CDRs) data, inclusive of 11,4B outgoing mobile calls of TIM, one of the major Italian telecommunication companies (30.8% of market share in Italy^a).

CDRs are collected for billing purposes by mobile network operators: more specifically, a CDR record of the user is created every time a phone interacts with the network, recording (i) the type of the event (incoming/outgoing call, transmission of a text message, consumption of a certain amount of data traffic), (ii) the pseudonym of the users involved (the one producing traffic and, eventually, e.g., in case of voice traffic, the other party involved), (iii) the timestamp of the event, and (iv) the cell network's antenna accessed for the event (i.e., to which the caller's phone was connected), that, to a wider extent, represents the location of the user [12, 13].

The CDRs of our dataset are limited to voice traffic and have been provided by TIM after some pre-processing steps. First of all, CDRs have been enriched with demographic data from the Customer Relations Management, in order to be able to represent users in terms of gender and age ranges. CDRs have then been filtered at 99% percentile on number of daily calls per user, in order to remove edge cases that are not representative of the general population (e.g., call centers). In particular, if the number of calls for a user during a day exceeds the threshold, all the CDRs associated with that user for that day are removed from the dataset. Finally, data have been aggregated by city, hour, gender and age-range, getting rid of the identities (even if already pseudoanonymized) of users. Thus, for each city and hour, the dataset contains: (i) the number of outgoing calls divided by gender, (ii) the number of outgoing calls divided by age range, and (iii) the total number of outgoing calls.

Regarding the identification of our cities, we have adopted the definition developed, in 2012, jointly by the European Commission and the Organization for Economic Co-operation and Development (OECD) [14]: a city is a local administrative unit (LAU) where the majority of the population lives in an urban centre of at least 50 000 inhabitants. The definition provides also a division of European cities into 6 size classes: S, M, L, XL, XXL and Global City. We have considered 76 Italian cities that fall into the OECD definition and group them in *Small* (S), *Medium* (M), *Large* (L, XL, XXL) ones. Notice that no city in Italy can be categorized, according to OECD definition, as Global City, since no Italian city has more than 5 million inhabitants.

Hence, if we define $calls_h(c, d)$ as the number of calls for a city c , during a day d and an hour h , the timeseries of the calls, or *city's activity pattern* (A), is a timeseries of the values $A_h(c, d)$ where

$$A_h(c, d) = \frac{calls_h(c, d)}{\sum_{h \in [0, 23]} calls_h(c, d)}.$$

It is worth highlighting that we are considering the percentage of calls over the day for each hour and city. Thus, we can compare different cities independently of the absolute number of outgoing calls.

Finally, we identify the following socio-economic indicators to investigate the economic role (i.e., city's wealth), the attractiveness for foreigners (e.g., immigrants), and the incoming and outgoing commuting patterns of the highly synchronized Italian cities:

- *Resident population*: The absolute number of the resident population in a city.^b
- *Foreign population*: The absolute number of the foreign-born population in a city.^b
- *Population density*: The ratio between the resident population and the city surface.^b
- *Foreign percentage*: The percentage of the foreign population over the resident population for a city.^b
- *Average income*: The average yearly income per tax payer^b.
- *In-out commuters ratio*: The ratio between commuters moving to a city X for work or study reasons and commuters moving from a city X for work or study reasons.^c
- *Incoming commuter ratio*: The ratio between commuters moving to a city X for work or study reasons and the resident population of that city.^c
- *Outgoing commuter ratio*: The ratio between commuters moving out from a city X for work or study reasons and the resident population of that city.^c

2.2 Dynamic time warping

In order to compute the synchronization between the activity patterns of each pair of our cities, we have used the Dynamic Time Warping (DTW) distance algorithm [15]. DTW has been extensively adopted in speech recognition [16], computer vision [17, 18], natural language processing [19, 20], and image matching and handwritten recognition [21] as a measure of similarity between timeseries. The algorithm provides an estimate of the optimal match between two timeseries, including possible compression, expansion or lags in sections of the sequences. For example, DTW can capture similarities in walking activities, even if an individual is walking faster than the other. Thus, DTW can remove the lag due to the circadian rhythms characterizing our timeseries [22, 23]. For this reason, it provides a more correct notion of similarity between cities' activity patterns than an approach based on sliding-window correlation [24].

More specifically, assuming two timeseries $X = (x_1, \dots, x_M)$ and $Y = (y_1, \dots, y_N)$ a DTW path $P = (p_1, \dots, p_K)$ is a sequence of tuples of indices where $p_k = (m_k, n_k) \in [1, \dots, M] \times [1, \dots, N]$ is subject to the following constraints:

1. $p_1 = (1, 1)$ and $p_K = (M, N)$
2. $m_1 \leq m_2 \leq \dots \leq m_K$ and $n_1 \leq n_2 \leq \dots \leq n_K$
3. $p_{k+1} - p_k \in \{(1, 0), (0, 1), (1, 1)\}$ for $k \in [1 : K - 1]$.

Given a distance function d (e.g., Euclidean distance), the cost of a path c_p is defined as $c_p(X, Y) = \sum_{k=1}^K d(x_{m_k}, y_{n_k})$. The DTW distance between X and Y is hence defined as the cost of the wrapping path p^* having minimal total cost among all the possible wrapping paths.

By considering the activity pattern timeseries associated with the activity level of a city, we have obtained the DTW distance between the timeseries of all the couples of cities for a given day. Hence, the higher the DTW distance between a couple of cities, the lower the synchronization of their activity pattern timeseries. Moreover, we have computed the mean and variance of the DTW distances, during weekdays and weekends, for each couple of cities. Mean and variance are estimated by using the jackknife resampling procedure [25].

In order to investigate the association between the DTW distances and the socio-economic indicators listed in Sect. 2.1, for each city we have considered the average of the means previously computed using the jackknife resampling method. Then, we have computed the variance-weighted average of the DTW distances for each city by using the inverse-variance weighting procedure [26]. This method permits aggregation of two or more random variables (i.e., DTW distances) to minimize the variance of the weighted average.

Finally, for each city the variance-weighted average of the DTW distances is associated to each of the socio-economic indicators by means of Spearman bivariate correlations. The Spearman bivariate correlation measures the strength and direction of the association between two variables. Specifically, the Spearman coefficient is a number between -1 and $+1$, where -1 means perfect negative correlation, $+1$ indicates perfect positive correlation and 0 indicates no correlation.

2.3 Bootstrap procedure

To obtain an accurate estimation of the variance of the parameters of our fits, we performed a bootstrap procedure. Bootstrap resampling method is a widely used technique to infer properties of an estimator by sampling the original data repeatedly. As such, we have performed a group bootstrap by extracting a city and adding to our bootstrap sample all the couples containing the extracted city. This procedure guarantees to preserve, at each bootstrap iteration, all the correlations that a city has with other cities, since no couples that include the selected city are left out. Our bootstrap procedure follows three steps:

- (i) For each group of n cities of the same size (*Large*, *Medium* and *Small*) extract n cities with replacement;
- (ii) Create the dataset with couples of cities for the bootstrap iteration using all the possible combinations of extracted cities (excluding the couples with the same city);
- (iii) Perform a Weighted Least-Square Regression (WLS) using as weights the variance previously computed using the jackknife sampling method.

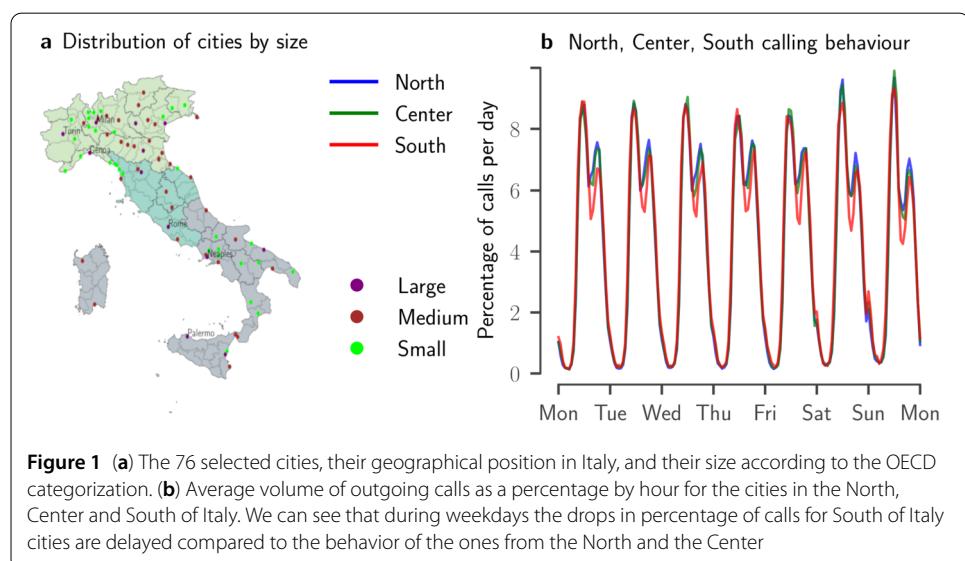
For each bootstrap iteration we implement a Weighted Least-Square Regression and collect the values of the slope m and the intercept q of the fit. Finally, obtained results were evaluated by performing a T-test to asses whether the slope differs from zero.

3 Results

The activity level of a city is the result of the combined behavioural patterns of different agents (i.e., individuals) and external constraints such as working schedules, school timetables or vacations. Such activity is mirrored by the number of calls placed in a city during a day: therefore, we have considered the percentage of outgoing calls per hour in each city (*city's activity timeseries*) as a proxy of the activity level of the city over time.

Our sample equally represents cities of classes *Medium* and *Small*, while we have less *Large* cities. Moreover, the cities we have considered are evenly spread across all regions of Italy, as can be seen in Fig. 1(a). Interestingly, by examining the activity timeseries associated with North, Center and South of Italy (see Fig. 1(b)), a similar pattern emerges for North and Center while South is characterized by a shift in the drop during lunchtime. For this reason, we have used DTW distance to compute the synchronization between *cities' activity timeseries*. Indeed, DTW distance removes the influence of the observed lags due to circadian rhythms (see Sect. 2.2).

In Fig. 2, each cell of the heatmap represents the mean value over weekdays of the DTW distance between two cities. The cities on both axes are ordered by total volume of calls from lowest (top left corner) to highest (bottom right corner). Two emerging clusters can be observed: one at the top right corner, where smaller cities with lower call volume are located, tends to have larger average distance between the communication activity timeseries; the other one can be observed in the bottom right corner, where large cities (with higher call volume) tend to have smaller mean DTW distance. The mean value of the DTW distance roughly increases when the volume of calls for a city decreases (i.e., it roughly scales with the size of the city): thus, the mean DTW distance for medium call volume cities is lower than the one for smaller cities, but higher than the DTW distance for larger cities.



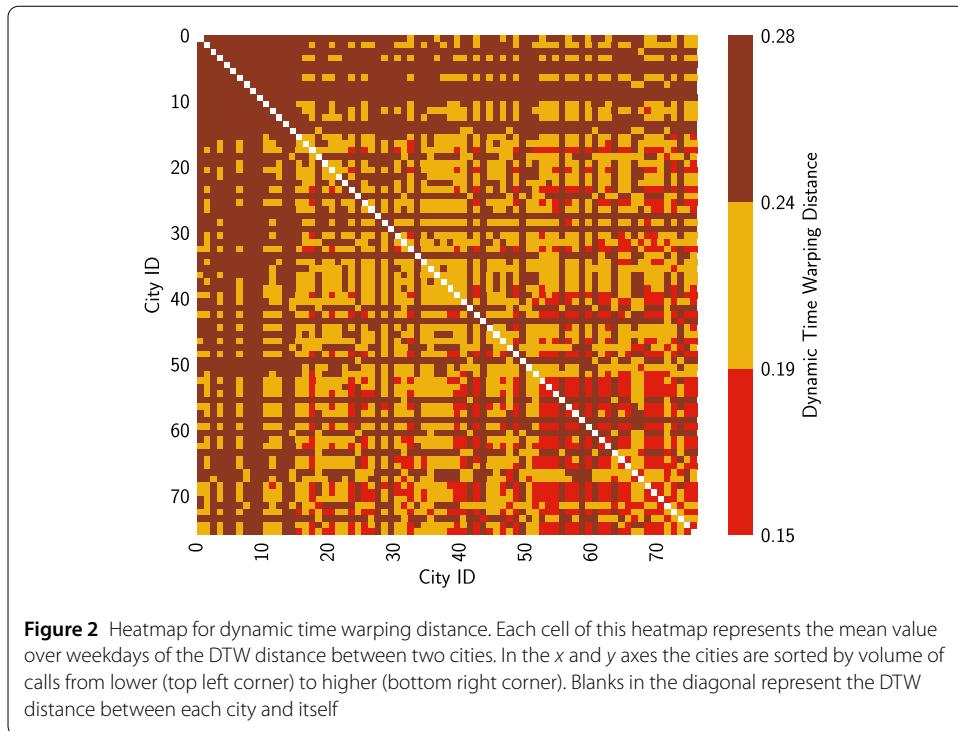


Figure 2 Heatmap for dynamic time warping distance. Each cell of this heatmap represents the mean value over weekdays of the DTW distance between two cities. In the x and y axes the cities are sorted by volume of calls from lower (top left corner) to higher (bottom right corner). Blanks in the diagonal represent the DTW distance between each city and itself

As previously said, we have performed a bootstrap for cities of the same size (*Large*, *Medium* and *Small*) and a WLS is fitted using the mean and the variance of the DTW distance between cities' activity pattern timeseries. As seen in Fig. 2, DTW distance roughly decreases when the number of outgoing calls increases and we can roughly divide the cities into three clusters based on DTW distance. Remarkably, this relationship still holds when considering the division of Italian cities into three size classes (*Large*, *Medium* and *Small*) according to OECD definition (see Sect. 2.1 for details). Cities of the same size appear to be more similar: two large cities (such as Turin and Milan) are more similar than two medium cities (such as Padua and Modena), and medium cities are more similar than small cities.

As we can see in Fig. 3, the similarity decreases with the increase of the distance between cities in a consistent way for all classes (see Table 1 for a complete report of the computed statistical measures) and the 95% confidence intervals for all classes are mostly not overlapping. Two large cities that are close to each other such as Milan and Turin are more similar than more distant large cities such as Milan and Rome. However, the similarity between large cities is higher than the similarity between medium cities, independently of the distance, i.e. Milan and Rome (two distant large cities) are more similar than Padua and Modena (two medium cities that are closer). The same relationship holds for medium and small cities: two medium cities are more similar to each other than two small cities, and the similarity decreases when spatial distance between cities increases.

In Fig. 4, during weekends, we can observe a similar pattern as for weekdays (see Table 2 for a complete report of the computed statistical measures). In particular, the ordering of the cities' classes is consistent with weekdays and the drift due to distance in the similarity is still visible, although confidence intervals are not as clearly separated as the weekdays' ones in Fig. 3.

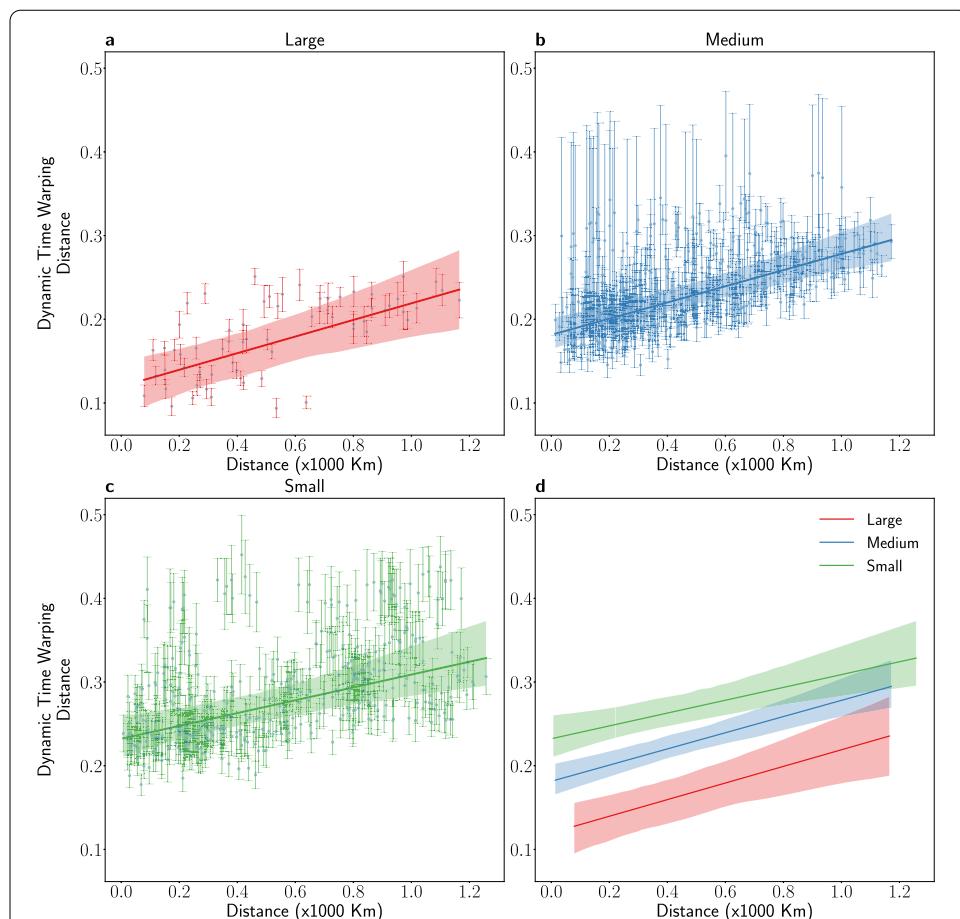


Figure 3 Bootstrap for weekdays. (a) shows the results of the bootstrap for couples of Large cities during the weekdays. The point and the bars represent the mean and the variance obtained for each couple through the jackknife sampling method. The shaded area represents the 95% confidence interval obtained using the bootstrap method described in *Materials and methods* and the line represents the bootstrapped Weighted Least-Square Regression (WLS) fit. (b) as (a) but for the Medium cities. (c) as (a) but for Small cities. (d) summarizes the estimates of the bootstrap regression previously obtained in figures (a)–(c). Note that 95% confidence intervals for all the classes mostly do not overlap. Detailed statistics are reported in Table 1

Table 1 Bootstrap results for weekdays. The table reports the mean values of intercept q and slope m for 500 bootstrap iterations. The bootstrapped 95% Confidence Intervals (CI) and the p -value p for the t-test on the slope m ($p \leq 0.05$: *, $p \leq 0.01$: **, $p \leq 0.001$: ****) are also reported

Size	q	m	p
Large	0.120 CI [0.084–0.151]	0.099 CI [0.043–0.153]	**
Medium	0.181 CI [0.164–0.201]	0.097 CI [0.068–0.131]	***
Small	0.232 CI [0.210–0.260]	0.077 CI [0.041–0.121]	***

Furthermore, Table 3 reports the Spearman correlation scores between the variance-weighted average of the DTW distances and several socio-economic indicators characterizing (i) the size of a city (i.e., resident population and population density), (ii) the city's attractiveness for foreigners and immigrants (i.e., foreign-born population and percentage of foreigners per resident population), (iii) the city's wealth (i.e., yearly average income per tax payer), and (iv) the outgoing and incoming commuting patterns.

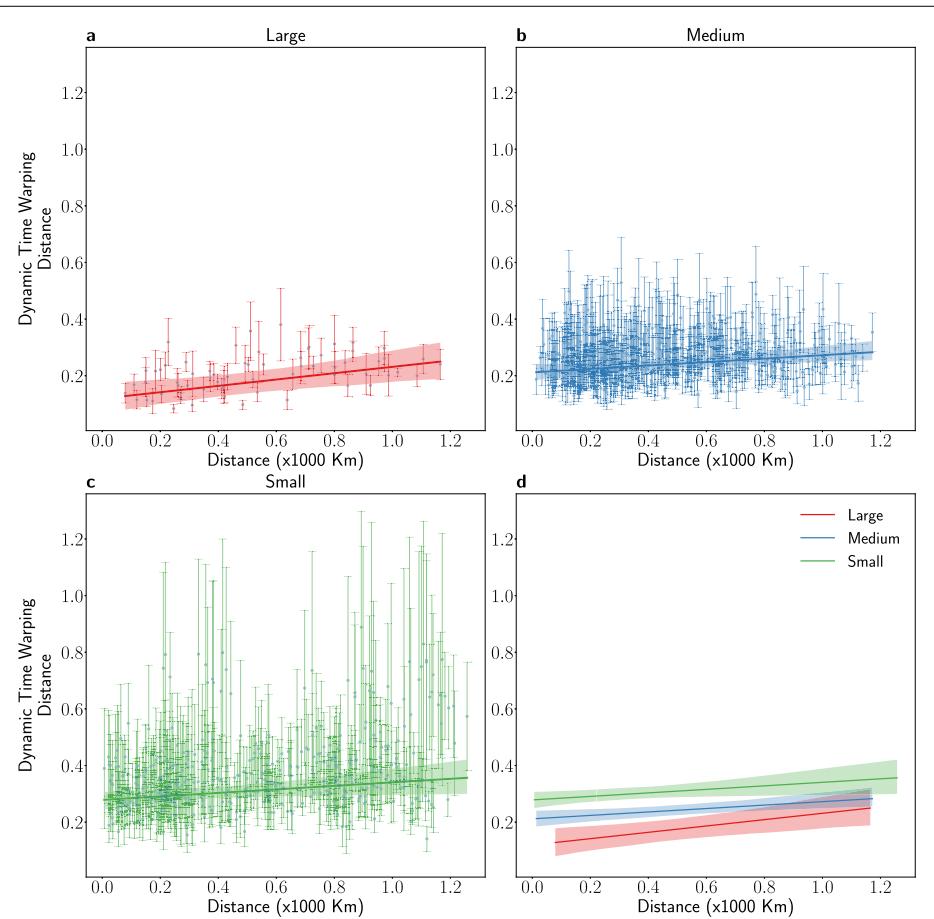


Figure 4 Bootstrap for weekends. (a) shows the results of the bootstrap for couples of Large cities during the weekends. The point and the bars represent the mean and the standard deviation obtained for each couple using the jackknife sampling method. The shaded area represents the 95% confidence interval obtained using the bootstrap method described in *Materials and methods* and the line represents the bootstrapped Weighted Least-Square Regression (WLS) fit. (b) as (a) but for the Medium cities. (c) as (a) but for Small cities. (d) summarizes the estimates of the bootstrap regression previously obtained in figures (a)–(c). Note that error bars and confidence intervals are larger, due to less data availability for weekends (our dataset consists of 6 weekend days and 18 weekdays)

Table 2 Bootstrap results for weekends. The table reports the mean values of intercept q and slope m for 500 bootstrap iterations. The bootstrapped 95% Confidence Intervals (CI) and the p -value p for the t-test on the slope m ($p \leq 0.05$: *, $p \leq 0.01$: **, $p \leq 0.001$: ***>) are also reported

Size	q	m	p
Large	0.119 CI [0.067–0.174]	0.112 CI [0.046–0.192]	*
Medium	0.212 CI [0.183–0.240]	0.061 CI [0.023–0.114]	**
Small	0.279 CI [0.249–0.307]	0.062 CI [0.011–0.131]	*

Our results show that the variance-weighted average of the DTW distances is negatively associated with the absolute number of resident population (Spearman's $\rho = -0.614^{***}$) and the population density (Spearman's $\rho = -0.477^{***}$) as well as with the absolute number of foreign-born population (Spearman's $\rho = -0.653^{***}$) and the foreign percentage (Spearman's $\rho = -0.409^{***}$). Thus, more synchronized cities seem to be both more populated and more attractive for foreigners (i.e., tourists, immigrants). Interestingly, the variance-

Table 3 Spearman correlation between the variance-weighted average of the DTW distances and (i) resident population (absolute number), (ii) foreign population (absolute number), (iii) population density, (iv) percentage of foreigners per resident population, (v) yearly average income per tax payer, (vi) ratio between commuters coming to a city and commuters leaving a city, (vii) percentage of incoming commuters per resident population, and (viii) percentage of outgoing commuters per resident population ($p \leq 0.05$: *, $p \leq 0.01$: **, $p \leq 0.001$: ***)

	DTW variance-weighted avg.
Resident population	-0.614***
Foreign population	-0.653***
Population density	-0.477***
Foreign percentage	-0.409***
Average income	-0.349**
In-out commuters ratio	-0.038
Incoming commuter ratio	-0.087
Outgoing commuter ratio	-0.236*

weighted average of the DTW distances also correlates negatively with yearly average income per tax payer (Spearman's $\rho = -0.349^{**}$), showing a relationship between more synchronized cities and the rich ones. We have also investigated the effect of *activity timeseries* synchronization on the commuting (incoming and outgoing) patterns of each city. As shown again in Table 3, we have not found significant correlations, with the exception of a slightly significant negative association (Spearman's $\rho = -0.236^*$) between the variance-weighted average of the DTW distances and the outgoing commuters' ratio (computed as the ratio between the number of outgoing commuters and the number of incoming + outgoing commuters).

We have also tested the correlation between the mean DTW distance and the spatial distance for each couple of cities. Our results (Spearman's $\rho = 0.205^{***}$) show that the increasing spatial distance is associated with a lower communication synchronization. Hence, it seems that regional economies are playing a role in the communication synchronization of our cities.

Finally, in Table 4 we report the 20 cities with the lowest variance-weighted average of the DTW distances, namely the cities with higher level of activity timeseries synchronization. Interestingly, Rome appears as the most synchronized city in Italy (mean DTW distance = 0.168) and this may be explained by a mixture of several factors such as its political and economic role (i.e., Rome is the national capital and the second wealthiest city in Italy), its size (i.e., Rome is the most populated city in Italy), and its attractiveness for foreigners (i.e., tourists). Other cities showing high degree of activity timeseries' synchronization are important cities for the maritime trade routes (i.e., Trieste and Genoa are the major Italian seaports for trade of goods and flows of people). Again, relevant touristic cities, such as Florence, Rimini, Ravenna, Venice, Verona, Milan are among the ones with higher levels of activity timeseries' synchronization. Finally, it is worth noting that only one city (Palermo) located in the South of Italy appears in the list of the 20 most synchronized ones.

4 Discussion

Cities' synchronization and similarity using mobile phone and social media (i.e., Twitter) data has been recently investigated in [27] and in [9]. In Grauwin *et al.* [27], three global cities (New York, London and Hong Kong) are studied by means of mobile phone usage patterns. The paper shows that these three large cities, despite the distance, have compa-

Table 4 The list of the 20 cities with the lowest variance-weighted average of the DTW distances. It is worth noting that cities with lower variance-weighted average DTW distances are the more synchronized ones

Ranking	City	DTW variance-weighted avg.
1	Rome	0.168
2	Genoa	0.183
3	Florence	0.186
4	Brescia	0.190
5	Turin	0.193
6	Trieste	0.197
7	Rimini	0.200
8	Ravenna	0.208
9	Bologna	0.210
10	Udine	0.211
11	Milan	0.217
12	Como	0.218
13	Venice	0.218
14	Verona	0.224
15	Terni	0.225
16	Bolzano	0.227
17	Modena	0.230
18	Bergamo	0.231
19	Ancona	0.231
20	Palermo	0.231

rable and common usage patterns, especially in the core business districts of the cities. In Morales *et al.* [9], an analysis of the synchronization of large world cities is presented using Twitter data. In this work, a cluster of similar large cities (Middle Eastern, European and African cities) is detected.

Our results, based on CDR data for 76 Italian cities, provide some evidence in support of these findings, at least for a subset of European cities (Italian large cities), by considering only the DTW distance. Indeed, we can observe, based on the outgoing calls' similarity patterns, an emerging cluster of similar Italian large cities (see Fig. 2). We further investigated the city's similarity and synchronization concept, by considering different scales of cities (Large, Medium and Small) and exploring how distance influences similarity between cities. After removing the effects of circadian rhythms by using DTW as timeseries distance measurement, we have analyzed the effect of spatial distance over cities' similarity.

We can suppose that Italian large cities act as *gateway cities* [11]: a gateway city is a city that plays the role of hub and central node for resources and capital circulation for the whole region where the city is located. *Gateway cities* host tertiary services such as banks, trading centers, headquarters of large companies that require a high degree of synchronization [28]. In our paper, we observed this behaviour for Italian large cities, that are more similar in rhythms to each other, despite the distance, than medium cities are to medium cities or small cities are to small cities. This is confirmed also by [29], that describes North Italy as a complex interconnected area (*city-region*), where larger cities provide advanced services, such as financial trading centers, for the whole area and act as a gateway for information and commercial flows for the whole region. Interestingly, in our study the 20 cities showing the highest degree of activity timeseries' synchronization are all located in the North (15 out of 20) or in the Center (4 out of 20) of Italy, with the exception of Palermo. Again, our results indicate that highly synchronized cities play a relevant

economic role (these cities have higher levels of average yearly income per tax payer) and are more attractive for foreign-born people (i.e. immigrants, tourists, etc.).

As shown in [28], cities follow a common development trajectory and after the population reaches a certain threshold (1.2 million people for the US case analyzed in [28]), the economical development path moves from primary industries (e.g., agriculture, mining, etc.) to tertiary industries such as banks and services. These kind of industries require a higher level of synchronization, such as the case of brokers trading stocks, or large industries that have headquarters spread all over Italy and the world, thus reducing the influence of distance over cities' similarity.

Differently for small cities, distance has a larger role in determining the communication synchronization and the similarity, since economy is more bounded to local productions and smaller industries. In the case of medium cities, as shown in [28], the cities are ongoing an industry transformation from primary to tertiary, hence the contribution of the distance over similarity is less than for the case of small cities but still bigger than the one for large cities.

These findings support also the theory exposed in [30], where, by using a scaling model and analyzing social and economical factors, such as GDP, wages, number of crimes, it is shown that large cities have a temporal self-similarity, in terms of higher and faster patterns of social interaction, walking speed of pedestrians, number of employed in research and development. It is also shown that smaller cities, when growing, follow a common social dynamic as the larger ones: when a city increases in population, it tends to accelerate its rhythms and have faster behaviours and technical innovations rates. This is confirmed by our observations, showing a scaling and division of cities based on city's size: large cities are more similar in rhythms to large cities, despite the distance, than medium cities are to medium cities, and medium cities are more similar to medium cities than small cities are to small cities.

However, our work has several limitations: for example, while we could hypothesize that the high synchronization of some cities (e.g. Trieste, Genoa, Ancona) is due to their importance for maritime trade routes, this result can not be validated by checking per-city correlations between the variance-weighted average of the DTW distances and the amount of traded goods (or similar indicators about commercial trades). Indeed, to the best of our knowledge, there is no available dataset that provides information about commercial trades at city-level granularity. Furthermore, even if our method could be applied for any city for which the CDR data is available, for our study we did not have access to any CDR data, for the same time period, for other world cities. Thus, we can not investigate the *gateway* role played by Italian cities worldwide.

5 Conclusion

In this paper, we have investigated Italian cities' communication synchronization and the influence of spatial distance over this communication synchronization by analyzing the CDR data of 76 Italian cities of different sizes. We found that larger cities tend to be more similar to each other than medium cities when considering only similarity between call patterns. We found also that similarity decreases when spatial distance between cities increases. The drift due to distance over similarity is less strong in large cities, that are gateway nodes for the Italian economical system, hence they have an emerging strongly connected and synchronized network, than for medium and small cities, that are more

bounded to local industries. We observed that the similarity decreases in a consistent way according to size of the city: large cities are more similar to large cities than medium cities are to medium cities, independently of the spatial distance, and the same holds for medium cities and small cities. Finally, our results have shown that cities with higher average synchronization tend to be richer and to attract more people from other places (e.g. tourists, business people, and immigrants).

Funding

Lorenzo Candeago is supported by a fellowship from TIM—Telecom Italia, SKIL (Semantic and Knowledge Innovation Lab) Joint Open Lab. The work of Bruno Lepri and Paolo Bosetti was conducted within the agreement between SDA Bocconi School of Management and Fondazione Bruno Kessler.

Abbreviations

CDR, Call Detail Records; CRM, Customer Relations Management; DTW, Dynamic Time Warping; LAU, Local Administrative Unit; OECD, Organization for Economic Co-operation and Development; WLS, Weighted Least-Square Regression.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Conceived the study: LC, PB, BL. Designed experiments and analyzed the data: LC, GB, PB. Wrote the paper: LC, PB, BL. Dataset preparation: MV. All authors read, reviewed and approved the final manuscript.

Author details

¹Fondazione Bruno Kessler, Trento, Italy. ²SKIL (Semantic and Knowledge Innovation Lab) Joint Open Lab, TIM—Telecom Italia, Trento, Italy. ³Department of Information Engineering and Computer Science, University of Trento, Trento, Italy.

⁴Department of Mathematics, University of Trento, Trento, Italy. ⁵SDA Bocconi School of Management, Milan, Italy.

Endnotes

^a <https://www.statista.com/statistics/710559/mobile-network-provider-market-share-italy/>

^b 15° Censimento generale della popolazione e delle abitazioni, ISTAT, 2011, <https://www.istat.it/it/censimenti-permanenti/censimenti-precedenti/popolazione-e-abitazioni/popolazione-2011>.

^c 15° Censimento generale della popolazione e delle abitazioni, Matrici del pendolarismo, ISTAT, 2011 <https://www.istat.it/it/archivio/139381>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 5 December 2018 Accepted: 20 May 2019 Published online: 29 May 2019

References

1. Strogatz S (2003) Sync: the emerging science of spontaneous order. Theia, New York
2. Schneidman E, Berry MJ II, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440(7087):1007
3. Van Dongen S, Backeljau T, Matthysen E, Dhondt AA (1997) Synchronization of hatching date with budburst of individual host trees (*Quercus robur*) in the winter moth (*Operophtera brumata*) and its fitness consequences. *J Anim Ecol* 66:113–121
4. Sumpter DJ (2010) Collective animal behavior. Princeton University Press, Princeton
5. Neda Z, Ravasz E, Brechet Y, Vicsek T, Barabasi A-L (2000) Self-organizing process: the sound of many hands clapping. *Nature* 403:849–850
6. Saavedra S, Hagerty K, Uzzi B (2011) Synchronicity, instant messaging, and performance among financial traders. *Proc Natl Acad Sci USA* 108:5296–5301
7. Mamei M, Pancotto F, De Nadai M, Lepri B, Vescovi M, Zambonelli F, Pentland A (2018) Is social capital associated with synchronization in human communication? An analysis of Italian call records and measures of civic engagement. *EPJ Data Sci* 7(1):25
8. Levinson D (2012) Network structure and city size. *PLoS ONE* 7(1):29721
9. Morales AJ, Avilalala V, Benito RM, Bar-Yam Y (2017) Global patterns of synchronization in human communications. *J R Soc Interface* 14(128):20161048
10. Sassen S (2004) The global city: introducing a concept. *Brown J World Aff* 11:27
11. Short JR, Breitbach C, Buckman S, Essex J (2000) From world cities to gateway cities: extending the boundaries of globalization theory. *City* 4(3):317–340
12. Blondel VD, Decuyper A, Krings G (2015) A survey of results on mobile phone datasets analysis. *EPJ Data Sci* 4:10
13. Barlacchi G, De Nadai M, Larcher R, Casella A, Chitic C, Torrisi G, Antonelli F, Vespiagnani A, Pentland A, Lepri B (2015) A multi-source dataset of urban life in the city of Milan and the Province of Trentino. *Sci Data* 2:150055
14. Dijkstra L, Poelman H (2012) Cities in Europe: the new OECD-EC definition. *Regional Focus* 1(2012):1–13
15. Müller M (2007) Dynamic time warping. In: Information retrieval for music and motion, pp 69–84

16. Gaikwad SK, Gawali BW, Yannawar P (2010) A review on speech recognition technique. *Int J Comput Appl* 10(3):16–24
17. Antonini G, Thiran J-P (2006) Counting pedestrians in video sequences using trajectory clustering. *IEEE Trans Circuits Syst Video Technol* 16(8):1008–1020
18. Chen TP, Haussecker H, Bovyrin A, Belenov R, Rodyushkin K, Kuranoc A, Eruhimov V (2005) Computer vision workload analysis: case study of video surveillance systems. *Intel Technol J* 9(2):109–118
19. Radinsky K, Agichtein E, Gabrilovich E, Markovitch S (2011) A word at a time: computing word relatedness using temporal semantic analysis. In: Proceedings of the 20th international conference on World Wide Web. ACM, New York, pp 337–346
20. Myers CS, Rabiner LR (1981) A comparative study of several dynamic time-warping algorithms for connected-word recognition. *Bell Syst Tech J* 60(7):1389–1409
21. Rath TM, Manmatha R (2003) Word image matching using dynamic time warping. In: Computer vision and pattern recognition, 2003. Proceedings. 2003 IEEE computer society conference on, vol 2. IEEE, New York
22. Aledavood T, Lopez E, Roberts SGB, Reed-Tsochas F, Moro E, Dunbar RI, Saramaki J (2015) Daily rhythms in mobile telephone communication. *PLoS ONE* 10:0138098
23. Monsivais D, Ghosh A, Bhattacharya K, Dunbar RI, Kaski K (2017) Tracking urban human activity from mobile phone calling patterns. *PLoS Comput Biol* 13(11):1005824
24. Bartolini I, Ciaccia P, Patella M (2005) Warp: accurate retrieval of shapes using phase and time warping distance. *IEEE Trans Pattern Anal Mach Intell* 27(1):142–147
25. Bissell A (1977) The jackknife. *Appl Stat* 4(1):55–64
26. Hartung J, Knapp G, Sinha BK (2011) Statistical meta-analysis with applications, vol 738. Wiley, New York
27. Grauwin S, Sobolevsky S, Moritz S, Gómez I, Ratti C (2015) Towards a comparative science of cities: using mobile traffic records in New York, London and Hong Kong. In: Computational approaches for urban environments. Springer, New York, pp 363–387
28. Hong I, Frank MR, Rahwan I, Jung W-S, Youn H (2018) A common trajectory recapitulated by urban economies. arXiv preprint. [arXiv:1810.08330](https://arxiv.org/abs/1810.08330)
29. Garavaglia L (2014) The distribution of advanced business services in northern Italy: towards a polycentric metropolis model? *Métropoles* 14
30. Bettencourt LM, Lobo J, West GB (2008) Why are large cities faster? Universal scaling and self-similarity in urban organization and dynamics. *Eur Phys J B* 63(3):285–293

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com